UNITED STATES PATENT AND TRADEMARK OFFICE

MN

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/670,681	09/25/2003	Eric D. Brill	MS305080.1/MSFTP475US	7563

27195 7590 07/17/2007

AMIN. TUROCY & CALVIN, LLP
24TH FLOOR, NATIONAL CITY CENTER
1900 EAST NINTH STREET
CLEVELAND, OH 44114

EXAMINER	
JEAN GILLES, JUDE	

ART UNIT	PAPER NUMBER
2143	

MAIL DATE	DELIVERY MODE
07/17/2007	PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Office Action Summary	Application No.	Applicant(s)	
	10/670,681	BRILL ET AL.	
	Examiner	Art Unit	
	Jude J. Jean-Gilles	2143	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *25 September 2003*.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1-116* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-116* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☒ The drawing(s) filed on *25 September 2003* is/are: a)☒ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☒ Information Disclosure Statement(s) (PTO/SB/08) Paper No(s)/Mail Date *05/19/2005, 01/18/2005, 05/13/2004*.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____ .
5) ☐ Notice of Informal Patent Application
6) ☐ Other: _____.

DETAILED ACTION

This office action is responsive to communication filed on 09/25/2003.

Information Disclosure Statement

1. The references listed on the Information Disclosure Statement submitted on 05/14/2002 have been considered by the examiner (see attached PTO-1449A).

Claim Rejections - 35 USC § 102

2. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

> A person shall be entitled to a patent unless -
>
> (e) the invention was described in (1) an application for patent, published under section 122(b), by another filed in the United States before the invention by the applicant for patent or (2) a patent granted on an application for patent by another filed in the United States before the invention by the applicant for patent, except that an international application filed under the treaty defined in section 351(a) shall have the effects for purposes of this subsection of an application filed in the United States only if the international application designated the United States and was published under Article 21(2) of such treaty in the English language.

3. **Claims 1-112** are rejected under 35 U.S.C. 102(e) as being anticipated by Bailey et al (Bailey), Pub. No. 20060167864 A1.

Regarding **claims 1-112**, Bailey discloses:

1. A data analysis system (fig. 1), comprising:

a first component that facilitates generation of a first data set related to web page information obtained via a communication system (fig. 1, item 120); and

a second component that coordinates a data set relating to web page information from at least one distributed resource which interacts with the communication system; the

second data set is utilized to refine the first data set (see abstract, fig. 1; *note that the web crowler (160) generates the data set through the Internet*; 0037-0040, 0052).

2. The system of claim 1, the first component comprising an internet web crawler (120).

3. The system of claim 1, the first component comprising an intranet web crawler (120; *the crawler is usable equally in the Internet, as well as an Intranet*).

4. The system of claim 1, the second component further utilized to optimize reception of data from the distributed resources (164).

5. The system of claim 1, the second component provides a scheduling function to control reception of the second data set from the at least one distributed resource (147).

6. The system of claim 1, the second component utilized to facilitate communication traffic reduction via the communication system by employing a proper set of weak indicator functions representative of the first data set (162).

7. The system of claim 6, the second component further utilized to randomly select and transmit a weak indicator function selected from the proper set of weak indicator functions to at least one of the distributed resources (160, 162, 164).

8. The system of claim 1, the second component further utilized to compare the first data set and the second data set to detect spoof data retrieved by the first component (*comparing spoof data with a web crawler is inherent in the art*).

9. The system of claim 1, the second component further utilized to generate status information about data related to the first data set; the status information transmitted to at least one distributed resource (fig. 5; 0070).

10. The system of claim 9, the status information comprising, at least in part, a freshness flag to indicate freshness of information related to the first data set (fig. 5; 0070).

11. The system of claim 9, the status information comprising, at least in part, a hash of contents of information related to the first data set (fig. 5; 0070, 0076).

12. The system of claim 9, the status information comprising, at least in part, a copy of information of the first data set (fig. 5; 0070, 0076).

13. The system of claim 1, the communication system comprising an internet (110, 120, 130).

14. The system of claim 1, the communication system comprising a world wide web

(110, 120, 130).

15. The system of claim 1, the communication system comprising an intranet (110, 120, 130).

16. The system of claim 15, the intranet comprising a local area network . (130).

17. The system of claim 15, the intranet comprising a wide area network (110, 120, 130).

18. The system of claim 1, the distributed resources comprising clients of a server (110, 120, 130).

19. The system of claim 1, the distributed resources comprising trusted entities interactive with the communication system and the second component (fig. 2, 5,.

20. The system of claim 1, the first data set comprising internet web page data (0043, 0070, 0087; fig. 1 & 2).

21. The system of claim 1, the first data set comprising intranet web page data (0043, 0070, 0087; fig. 1 & 2).

22. The system of claim 1, the second data set utilized to add additional data to the first data set; the additional data comprising data unknown to the first component (0043, 0070, 0087; fig. 1 & 2).

23. The system of claim 1, the second data set comprising, at least in part, a hash of contents of at least one web page (0040, 0070, 0087; fig. 1, 2, & 5).

24. The system of claim 1, the second data set comprising, at least in part, a Uniform Resource Locator (URL) of at least one web page (0040, 0070, 0087; fig. 1, 2 & 5).

25. The system of claim 1, the second data set comprising, at least in part, a time stamp relating to an acquisition time for information about at least one web page (0043, 0070, 0087; fig. 1 & 2).

26. The system of claim 1, the second data set comprising, at least in part, a delta indication of changes to contents of at least one web page (0043, 0070, 0087; fig. 1 & 2).

27. The system of claim 26, the delta indication including, at least in part, a hash of previous contents of a web page and a hash of recent contents of the web page (0028, 0043, 0070, 0087; 0076, fig. 1 & 2).

28. The system of claim 1, the second data set comprising, at least in part, a status indication of changes to contents of at least one web page(0028, 0043, 0070, 0087; 0076, fig. 1 & 2).

29. The system of claim 28, the status indication including, at least in part, a percentage relating to an amount of change of contents of a web page (0028, 0043, 0070, 0087; 0076, fig. 1 & 2).

30. The system of claim 28, the status indication including, at least in part, a significance indicator to signify importance of changes in contents of a web page (0028, 0043, 0070, 0087; 0076, fig. 1 & 2).

31. The system of claim 1, the second data set comprising internet web page data (0028, 0043, 0070, 0087; 0076, fig. 1 & 2).

32. The system of claim 1, the second data set comprising intranet web page data (0028, 0043, 0070, 0087; 0076, fig. 1 & 2).

33. The system of claim 1, the second data set comprising data compiled utilizing at least one weak indicator function randomly selected from a set of weak indicator functions; the set of weak indicator functions representative of the first data set (0028, 0043, 0070, 0087; 0076, fig. 1 & 2).

34. The system of claim 1, further comprising a search component to accept at least one search query and generate at least one search reply having at least a portion of the first data set represented by information embedded in the search reply (0028, 0043, 0070, 0087; 0076, fig. 1 & 2).

35. The system of claim 1, further comprising a web page server component to construct web pages having at least a portion of the first data set represented by information embedded in at least one link found on at least one constructed web page (0028, 0043, 0070, 0087; 0076, fig. 1 & 2).

36. The system of claim 1 further comprising a storage component to store the first data set (0028, 0043, 0070, 0087; 0076, fig. 1 & 2).

37. A method for facilitating data analysis, comprising:
generating a first data set relating to a second data set obtained from web pages interactive with a communication system (see abstract; fig. 1; (0037-0040, 0052); receiving a third data set from at least one distributed resource that is interactive with the communication system; the third data set comprising web page related information generated by the distributed resource; and refining the second data set to reflect information obtained from the third data set (0084-0088).

38. The method of claim 37, the first data set comprising a representation of the second data set (see abstract; fig. 1; (0037-0040, 0052).

39. The method of claim 38, the representation of the second data set comprising, at least in part, a hash of contents of at least one web page contained in the second data set (0028, 0043, 0070, 0087; 0076, fig. 1 & 2).

40. The method of claim 38, the representation of the second data set comprising, at least in part, a status indication of at least one web page contained in the second data set (0028, 0043, 0070, 0087; 0076, fig. 1 & 2).

41. The method of claim 40, the status indication comprising a freshness flag to indicate if the web page information is current (fig. 5; 0070).

42. The method of claim 37, the first data set comprising a copy of the second data set (0028, 0043, 0070, 0087; 0076, fig. 1 & 2).

43. The method of claim 37, the second data set comprising web page information compiled by a web crawler (0028, 0043, 0070, 0087; 0076, fig. 1 & 2).

44. The method of claim 37, the third data set comprising web page information based upon client accessed web page information on the communication system.

45. The method of claim 37, the distributed resource comprising a client of a distributed crawler system (0028, 0043, 0070, 0087; 0076, fig. 1 & 2).

46. The method of claim 37, the communication system comprising an internet (fig. 1).

47. The method of claim 37, the communication system comprising an intranet (fig. 1).

48. The method of claim 37, refining the second data set comprising: adding unknown information to the second data set when new information is received from the distributed source via the third data set; updating existing information in the second data set when changes have occurred as indicated by the third data set; and resetting any indicators utilized to pass status information to the distributed resources after information from the third data set has been analyzed (0028, 0084-0088; 0037-0043, 0070, 0087; 0076, fig. 1 & 2).

49. The method of claim 37, further including: transmitting the first data set to at least one distributed resource that is interactive with the communication system making the first data set available to be utilized by the distributed resource to generate the third data set (0028, 0084-0088; 0037-0043, 0070, 0087; 0076, fig. 1 & 2).

50. The method of claim 38, further including: generating a set of weak indicator

functions to represent the second data set; and selecting random weak indicator functions from the set of weak indicator functions to transmit to the distributed resources as the first data set (0028, 0084-0088; 0037-0043, 0070, 0087; 0076, fig. 1 & 2).

51. The method of claim 50, the set of weak indicator functions comprising a proper set of weak indicator functions such that a non-zero probability exists that a randomly selected weak indicator function can identify a new web page (0028, 0084-0088; 0037-0043, 0070, 0087; 0076, fig. 1 & 2).

52. The method of claim 50, generating a set of weak indicator functions comprising: providing a dictionary representative of the second data set; partitioning randomly the dictionary into non-overlapping subdictionaries; and creating a function where I(x)=1 if and only if at least one subdictionary's weak indicator function is equal to one (0076-0080).

53. The method of claim 37, further including: comparing the third data set to the second data set to reveal spoof data included in the second data set (0028, 0084-0088; 0037-0043, 0070, 0087; 0076, fig. 1 & 2).

54. The method of claim 37, further including: optimizing reception of at least one third data set through scheduling of the distributed resources (0028, 0084-0088; 0037-0043, 0070, 0087; 0076, fig. 1 & 2).

55. The method of claim 37, further including: receiving a web page search query from at least one distributed resource; generating a web search results page in response to the web page search query from the distributed resource; embedding portions of the first data set in links found on the web search results page; and transmitting the web search results page as a representation of at least a portion of the second data set to the distributed resource (0028, 0084-0088; 0037-0043, 0070, 0087; 0076, fig. 1 & 2).

56. The method of claim 37, further including: constructing a web page utilizing at least a portion of the first data set to embed information about links found in the web page; and transmitting the web page to disseminate the first data set to at least one distributed resource (0028, 0084-0088; 0037-0043, 0070, 0087; 0076, fig. 1 & 2).

57. A data analysis system, comprising: means for generating at least one first data set from a communication system; means for receiving and coordinating at least one second data set from at least one distributed resource which interacts with the communication system; and means for refining the first data set utilizing at least one second data set (0028, 0084-0088; 0037-0043, 0070, 0087; 0076, fig. 1 & 2).

61. A data analysis system, comprising: a first component that generates web page information from at least one visited web site for utilization in a distributed web crawling

system; the web page information transmitted by the first component to a second component via a communication system (0028, 0084-0088; 0037-0043, 0070, 0087; 0076, fig. 1 & 2).

92. A method for facilitating data analysis, comprising: compiling a first data set derived from accessing web pages via a communication system; and transmitting, selectively, the first data set to an entity of a distributed crawling system that is interactive with the communication system (0028, 0084-0088; 0037-0043, 0070, 0087; 0076, fig. 1 & 2).

113. A data packet transmitted between two or more computer components that facilitate information gathering, the data packet comprising, at least in part, information relating to web crawling that utilizes, at least in part, a distributed system for gathering information about web pages (0028, 0084-0088; 0037-0043, 0070, 0087; 0076, fig. 1 & 2).

Claims 58-60, 62-91, 93-112, and 114-116 are similar to other claims addressed above (see rejection of claims 2-56 above).

Conclusion

4. Any inquiry concerning this communication or earlier communications from examiner should be directed to Jude Jean-Gilles whose telephone number is (571) 272-3914. The examiner can normally be reached on Monday-Thursday and every other Friday from 8:00 AM to 5:30 PM.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, David Wiley, can be reached on (571) 272-3923. The fax phone number for the organization where this application or proceeding is assigned is (703) 305-3719.

Any inquiry of a general nature or relating to the status of this application or proceeding should be directed to the receptionist whose telephone number is (703) 305-3900.

Jude Jean-Gilles

Patent Examiner

Art Unit 2143

JJG

July 7, 2007

DAVID WILEY
SUPERVISORY PATENT EXAMINER
TECHNOLOGY CENTER 2100